

DOCUMENT RESUME

ED 295 695

JC 880 260

AUTHOR Mann, Rebecca
TITLE Measuring Writing Competency.
PUB DATE 26 Apr 88
NOTE 18p.; Paper presented at the Southeastern Conference on English in the Two-Year College (Louisville, KY, February 18-20, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Guides - Classroom Use - Guides (For Teachers) (052)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Evaluation Criteria; *Evaluation Methods; Holistic Evaluation; Test Reliability; *Writing Evaluation; *Writing Skills

ABSTRACT

In response to growing concern about the lack of basic writing skills, this paper presents an overview of the issues involved in selecting a method for the assessment of students' writing skills. After general criteria for determining the appropriateness of a writing evaluation procedure are outlined, the merits and limitations of objective tests and direct evaluation of writing samples are presented. While acknowledging the speed and ease with which objective tests can be administered, the paper emphasizes that only a writing sample permits a full assessment of writing ability and warns of the lack of correlation between direct and indirect measures. Next, a review is presented of the advantages and disadvantages of three scoring options for writing samples: (1) analytic scoring, which judges papers in terms of ideas, mechanics, organization, wording, and flavor; (2) holistic scoring, which responds to the writing sample as a whole; and (3) primary-trait scoring, a form of holistic scoring which considers the writing sample as a particular type of discourse and rates it accordingly. Finally, the paper discusses the political aspects of beginning a testing program, pointing out the need for faculty support of the program, for their willingness to participate as readers, and for consensus on the choice of a scoring option. Concluding comments argue that while no method of measurement is perfect, faculty have a responsibility to educate themselves and the administration about the scoring options available and to influence their institutions to utilize direct measures of writing ability. (AYC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED295695

MEASURING WRITING COMPETENCY

Rebecca Mann

April 26, 1988

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

R. S. Mann

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

Measuring Writing Competency

Articles concerning evaluation of competency of writing skills are appearing currently in the journals of English and writing. The general public has become more aware of the lack of basic writing skills of students through widespread reporting of a general decline in scores on objective tests. In search of higher levels of accountability for instruction, many legislatures are mandating minimum competency testing in basic skills at various levels of education, including higher education. Students face measurement of competency typically as part of placement or as an exit criteria from a course. Even accrediting agencies are demanding that institutions demonstrate that improvement results from instruction. At some point in the career of the English professional, his institution is, in all likelihood, going to be asked to defend its method of assessing writing skills. Even if faculty do not have to defend methods of assessment, they need to be informed of issues in writing assessment for improvement of classroom grading.

As Odell and Cooper have determined, the following should be investigated for any evaluation procedure used: 1) the underlying assumptions, 2) the extent to which these assumptions are consistent with current discourse theory, and 3) its limitations (43). Glick and Cole advise faculty to ask the practical questions in matching a writing test to a program: What information do I want? What is the purpose of the test? What are my resources in money and time? (5) With these guidelines in mind, faculty can begin to seek the best method for measuring writing competency.

The issues surrounding evaluation of writing skills might be less difficult to settle if an individual were allowed to make such decisions independently. The settlement of evaluation issues, however, is not usually a simple process, for faculty often are asked to comply with a decision of a group of individuals. At the postsecondary level the teacher only administers the test which is often chosen or made by a test development committee on the basis of cost effectiveness (Greenberg, "Competency" 367). When a committee determines test procedure primarily on cost effectiveness, it may decide to use indirect measures of writing ability, or objective tests, rather than direct measures, or writing samples. Citing increasing pressure from competency-based education adherents and the ease and speed of scoring objective tests, one director of freshman English chose to devise an objective test as an exit exam (Matalene 369), as have others faced with similar constraints. Greenberg explains that most research on evaluation of writing has been done by the College Entrance Examination Board, an affiliate of Educational Testing Service ("Competency" 368). Much of it examines indirect measures or multiple-choice tests which require no writing. Most composition teachers, Greenberg believes, do not accept indirect measures as valid tests of writing ability (368); therefore, as Greenberg asserts, classroom teachers themselves should develop writing tests to make certain that tests meet basic criteria and achieve their purpose (367).

Although objective tests fill a need by providing ease and speed of scoring, they do not allow full assessment of writing ability. Only a direct measure or writing sample in which the student is asked to invent and make choices in diction, syntax and content provides for full assessment. While objective tests can demonstrate greater reliability (the

degree to which a test measures what it is intended to measure) among members of the English profession. Further, these tests send a message to students that writing is not important (Matalene 369).

The debate of professionals in English over whether to use objective, standardized tests or direct assessment is a continuing one. Moran and Lunsford in Research in Composition and Rhetoric (1984) point out that a positive correlation between scores from multiple-choice usage tests and writing samples can result if the samples are read for the same features. But normally writing samples are read for content, and the correlation of results does not exist (214). The sizable disadvantages of direct assessment are costs in time and money. Edward M. White writes that "When concerned faculty have marshaled the evidence of need for--and shown themselves ready to take responsibility for--a writing testing program that supports the teaching of writing, they have found more support than they ever imagined" (Teaching xiv). In spite of White's encouraging words, faculty often have difficulty gathering the support of uninformed colleagues and administration to begin direct writing assessment.

Another disadvantage of direct assessment is that the response of the writer to the prompt may not include a wide enough range of skills expected for measurement, a factor the evaluator cannot control. On the other hand, indirect assessment requires no creation of prose at all. Examinees are asked to select or identify constructions that may never appear in their own prose (Moran and Lunsford 214). The total lack of production of writing with objective tests which purport to measure writing ability seems to be the reason that most English professionals believe that measures of writing competency should require a written sample.

Once the decision is made to use a direct writing measure, teachers should become familiar with scoring options. The three most often used are holistic, analytic, and primary-trait (Glick and Cole 5). Rather than evolve from a commonly held set of assumptions about evaluation, each method rests on its own set of assumptions. In fact, evaluation of writing proceeds without a common set of principles (Gere 44). A responsible faculty will investigate scoring options, weigh the advantages and disadvantages of each, and select the one that best fits the criteria and purpose of the program. In selecting the scoring option, faculty should keep in mind that use of writing samples necessitates obtaining at least two samples of more than one kind of task under circumstances that allow for editing and revising, since a person's writing will vary from day to day (Odell 118). An examination of the three most often used scoring options may be helpful as teachers decide on one that best suits their purpose.

The first of those scoring options, the analytic scale as developed by Paul B. Diederich, utilizes several criteria in judging papers. These criteria were determined by sixty judges from academic and non-academic fields as part of a study in factor analysis done in 1961 at Educational Testing Service (Diederich 5); they are ideas, mechanics, organization, wording, and flavor (9). Flavor refers to the characteristic which the judges themselves called style, but the researchers, including Diederich, renamed the trait flavor to indicate that quality which reveals the personality of the author (8). Each specific factor to be assessed is rated low, middle, or high on level of merit for that trait or quality of writing (55). Analytic scales may be constructed for assessment of various

kinds of writing, but the process is a time-consuming one and is less cost-efficient when applied to a large number of papers (Moran and Lunsford 221). In addition, Cooper points out that the analytic scale has a drawback of not being "sensitive to the variations in purpose, speaker role, and conception of audience which can occur in pieces written in the same mode" (14).

An illustration of the scale is given below. The doubled values reflect emphases in the writing program that first used the scale (Cooper 8).

Topic _____	Reader _____		Paper _____		
	Low		Middle		High
Ideas	2	4	6	8	10
Organization	2	4	6	8	10
Wording	1	2	3	4	5
Flavor	1	2	3	4	5 _____
Usage	1	2	3	4	5
Punctuation	1	2	3	4	5
Spelling	1	2	3	4	5
Handwriting	1	2	3	4	5 _____
Sum					_____

(Diederich 54)

Edward White discusses analytic scoring in Teaching and Assessing Writing (1985), adding to one's reasons to choose some other scoring option. White sees a theoretic problem with analytic scales:

There is no evidence that writing quality is the result of the accumulation of a series of subskills. To the contrary, the lack of agreement on subskills in the profession suggests that writing remains more than the sum of its parts and that the analytic theory that seeks to define and add up the subskills is fundamentally flawed. Analytic scoring is uneconomical, unreliable, pedagogically uncertain or destructive and theoretically bankrupt. (123-124).

White's statements are based on the premise that there is little agreement on the definitions or importance of various subskills. As a result of this lack of agreement, inter-rater reliability tends to be low (122). Unless faculty have adequate time to train themselves to the rating scale of high, medium, and low on each trait and are willing to accept the questionable theoretical basis of analytic scoring, this option might well be avoided.

Another direct measure of writing ability is holistic scoring. It differs from analytic scoring in a dramatic way. Instead of assessing selected composition features, it responds to student writing as a unit. Developed under the auspices of Educational Testing Service, holistic scoring is quick and practical as well as cost effective (White, "Holisticism" 402). Most important, however, is the fact that this scoring option permits the rating of total effectiveness of the writing sample, not just of certain features.

As in the analytic scale, the readers are given a set of instructions to guide their reading. Often called a "rubric," it distinguishes qualities of writing along a continuum that is worked out in advance of the scoring (Cooper 12). The score, usually from 1-4 or 1-6, indicates a rank-order fitting within a range for that assignment or sample. In fact, holistic scoring provides for a series of constraints that have contributed to the achievement of high reliability: 1) controlled essay reading; 2) scoring criteria guide; 3) sample papers; 4) checks on reading in progress;

5) multiple independent scoring; 6) evaluation and record keeping (White, "Holisticism" 404-405). Even though these high reliabilities may be overstated, as White believes, holistic scoring is the most successful means of scoring writing in quantity now available (408).

Several advantages of the holistic scoring option can be articulated, the most obvious being that it requires a response to the writing as a whole. In addition, it assumes a positive attitude by rewarding the writing for what is done well. Further, the careful definition of points on a scale allows for criterion-referencing (White, Teaching 124-125). The use of an even-numbered scale, usually the six-point scale, forces a choice of top-half or bottom-half on the rubric so that fewer students can be placed in a middle or in-between ranking on papers that may seem neither clearly good nor clearly weak. Moran and Lunsford recognize the time-efficiency factor of holistic scoring when a large number of papers need to be scored in a relatively short time, a clear advantage when rank-ordering students for placement, grouping, or exempting (223). White sums up the appeal and level of acceptance of holistic scoring in this statement: "In one decade, in a notoriously conservative and slow-moving profession, a new concept in testing and in teaching writing became accepted while no one was watching" (Teaching 19).

Even so, a skeptical stance should be maintained toward all testing, including holistic scoring. While holistic scoring does provide a general rating of a writing sample, it fails to isolate the criteria responsible for that rating. This failure would limit holistic scoring's use in teaching situations where diagnostic data is required. During research of the practice of evaluating assessment essays, Barritt, Stock, and Clark

have discovered another possible fault in holistic scoring. They found that two different readers can disagree on a writer's intended meaning as they re-create this meaning and decide whether, for example, the writer is serious or ironic. Consistent scores, they reason, may not always be appropriate scores. Doubt is cast on a basic tenet of holistic scoring: that two separate rankings of an essay with the same points or adjacent number of points are reliable rankings. This research concludes that high reliability with no validity is a possibility on some rankings of samples (324).

The demands for measurement of writing skills expressed by the American public--experts, legislators, and concerned citizens--reflect a growing concern over students' competence as writers. In an effort to meet these demands, the holistic scoring option, a quick and easy-to-learn one, has been devised. It can be used to measure competence, which has been defined as "the ability to discover what one wishes to say and to convey one's message through language, syntax, and content that are appropriate for one's audience and purpose" (Odell 103). While holistic scoring seems to provide an accepted scoring option for conveniently ranking writing samples, one must look further to find an option that can be adapted for instruction, focusing on features that are relevant to the kind of discourse the writing is, permitting precision in stating hypotheses and in devising criterion measures for research and curriculum development. Such an option is provided in primary trait scoring (Cooper 11) as developed by the National Assessment of Educational Progress.

Primary trait scoring is a type of holistic test that isolates a particular type of discourse and rates a writing sample as an example of its class. The main steps are "to define the universe of discourse, to

devise exercises which sample that universe precisely, to ensure cooperation of the writers, to devise workable scoring guides, and to use the guides" (Lloyd-Jones 37). The scoring guide delineates the characteristics necessary for success with a given rhetorical task. Gere has found the theoretical basis of primary trait scoring to be sound, appearing to accommodate communication intention with its foundation built on a three-part model which includes subject, discourse, and audience (48). Raymond has similarly judged the theoretical basis of this scoring option:

Primary trait scoring is receiving much attention today not only because its research design is sound but because its rhetorical assumptions are well articulated: instead of evaluating student writing in a vacuum, it evaluates that writing as a response to a specific task with a specific purpose communicated to the writer through carefully written instructions. (402)

Conceptually similar to holistic scoring, primary trait scoring focuses on some features of writing that holistic scoring ignores. White recommends this method for making judgments more narrow than general ability in writing, such as those needed in program evaluation, research, and teaching situations (Teaching 148). As an instructional tool, primary trait scoring can help identify performance levels for writing skills, including rhetorical and other features such as speaker role, audience and purpose. As a research tool, it sharpens the focus on complex skills required for a particular task and allows the rater to identify strengths and weaknesses precisely. Further, it provides for focus on one issue at a time as mastery of that particular trait is developed in a classroom setting (Moran and Lunsford 224-225).

Time and money required to isolate the primary traits that each mode must possess to be effective are obvious disadvantages. First, preliminary trials are essential to ensure that writers will understand the intent of

the assignment well enough to produce the expected kind of discourse (Lloyd-Jones 43). Second, developing a scoring guide is quite a labor that requires sixty to eighty hours of professional time. The list of what the guide includes is long:

A scoring guide consists of (1) the exercise itself, (2) a statement of the primary rhetorical trait of the writing which should be elicited by the exercise (a kind of statement of the limited test objective), (3) an interpretation of the exercise indicating how each element of the stimulus is presumed to affect the respondent (a kind of hypothesis about performance), (4) an interpretation of how the situation of the exercise is related to the posited primary trait (a synthesis of #2 and #3), (5) a system for defining the shorthand which is to be used in reporting descriptions of the writing (the actual scoring guide), (6) samples of papers which have been scored (definition of score points), and (7) discussions of why each sample paper was scored as it was (extensions of the definitions). (45)

The amount of time required to develop detailed scoring guides would eliminate primary trait scoring as an option for many teachers in the profession. However, the quality and clarity of the final product may justify the time spent. The scoring guide allows a response to the student that explains clearly why he has been scored high or low. The scoring guide becomes not only an instructional tool but also a credible standard for scoring. This standard, which remains the same from reader to reader, provides the high degree of validity and reliability inherent in primary trait scoring.

An examination of the three most often used scoring options for evaluation of competency of writing skills reinforces the fact that no ideal test exists. Further, writing tests can be devised to measure whatever one determines to be the essential criteria of the writing program. Raymond summarizes the frustration of measuring writing competency when he writes "...despite undeniable advances in composition,

there is still no consensus about what good writing is, much less about how it might be evoked or measured" (402). The practice of good pedagogy does not afford one the luxury of giving up in frustration. Public demands, if not a sense of professionalism, motivate the search for measurement techniques. If a teacher does not make the effort to know something about measurement, problems of unclear assignments, arbitrary grading, and insufficient attention to the writing process result (White, "Centrality" 3). Even minimal experience in the classroom reveals that measurement is central to the teaching of writing.

Finally, Kenneth Bruffee in Writing Assessment: Issues and Strategies (1986) discusses the political nature of the beginning of a testing program. Decisions about how an institution will educate and test its students must be made by vote or consensus among the faculty. Unless such consensus can be reached, he writes, the institution should not have a testing program. If a decision is reached to have one, the matter should be turned over to the faculty who are trained to deal with ignorance, their own as well as others' (96-97).

That testing is a political issue was evident at the recent Conference on College Composition and Communication Convention. Professor Renee Betz, Central Missouri State University, reported that she had liked the idea of using direct assessment even though her work load increased after her institution decided to use portfolio assessment techniques. However, many of her colleagues who had not read research supporting the need to have direct assessment resisted use of the method. Professor Betz found that everyone must agree on assessment for it to be a happy experience (Renee Betz). One way to keep direct assessment a happy experience, according to Professor Norman Betz, also of Central Missouri State University, is to

think about what kind of testing to use before a legislature mandates exit testing. He also reminded those present that once exit testing exists, the institution has the responsibility to provide remedial education for those who do not pass (Norman Betz).

A proven method of gathering the support of one's colleagues is the early involvement of faculty in planning stages of testing, according to Mary Fowles of Educational Testing Service. She advised that the research and reasons for using a procedure such as holistic scoring can come later after political support is assured (Fowles). Establishing the political groundwork for a viable testing program must proceed carefully to gain the support of the faculty.

Karen Greenberg, who oversees an exemplary program of direct writing assessment at Hunter College, City University of New York, agrees that teachers, not some outside group, should be readers of writing samples. In fact, at her institution, faculty contracts require being a reader. Her experience has shown that faculty resistance to serving as readers is reduced when faculty are involved in setting up the test (Greenberg, Conference).

Faculty should put aside local politics in the search for an agreed upon scoring option. The message, as experts from across the nation share their experience, is urgent: "We need to close the gap between teaching and assessing. If we don't do it, someone else is going to do it" (Brown). The fear is that uninformed legislators and testing committees will impose testing and scoring options on students and faculty that will adversely affect the teaching-learning process. Most faculty would agree with the statement that "Testing drives the curriculum" (Norman Betz). If writing teachers will develop the test themselves, they will maintain control over

their own discipline and profession. As a result, the issue of who should select the test to be used will never need to be raised. Faculty will be prepared to present the best option for their program, and outsiders will be assured that professionals are able to determine assessment techniques and scoring options for use at their institutions.

An individual faculty member can serve as a catalyst in a faculty seeking the best testing methods to measure competency in writing ability if he is informed of the advantages and disadvantages of the various scoring options available to use. Analytic scoring, similar to the kind of grading composition teachers traditionally use in the classroom, has been determined to be pedagogically unsound. Holistic scoring, while becoming widely accepted for its use in scoring large numbers of papers, requires reader training and a willingness on the part of the faculty to serve as readers. Primary trait scoring demands careful, detailed work to devise theoretically sound scoring guides that can serve instructional and diagnostic purposes. In choosing one of these three scoring options, a faculty needs to consider how much time can be spent reading samples or devising scoring guides. If a faculty determines that the benefits of testing with writing samples or direct measures do not outweigh the disadvantages, it will probably choose to test by means of indirect measure or objective tests even if it believes that these methods are not an accurate measure of writing ability. The costs in money and time often overshadow what a knowledgeable professional believes to be best for students in the writing program.

Leadership and skill in persuading faculty that the time spent in direct assessment is well spent befall the director of the program or

department head. In addition, that individual may find himself in the position of convincing a dean that the faculty needs the support of the institution to accomplish the goal of direct assessment through provision of the funds needed to increase salaries or to provide a bonus for readers. In all likelihood, an administrator will not make a decision to begin a direct assessment program unless the faculty requests it. The faculty will not request it unless each member is willing to spend the time implementing it. Only a true leader can persuade each constituency to make changes in the direction of direct assessment. An interview with one department head at a community college revealed that she believes that the decision to move to direct assessment must come from within the institution, not from the "top down." At the same time, faculty who are requesting such a testing program will need institutional support in terms of pay, time, and regulations or policies that reinforce the program, such as closer adherence to the drop/add policies, especially when writing samples are taken the first week of class (Barber).

In conclusion, while no method of measurement is perfect, faculty can educate themselves on scoring options that are available. They can read the recent research to understand underlying assumptions of scoring options and discourse theory in order to determine the limitations and uses of available techniques. They can discuss issues of assessment among themselves and with administration to prevent the making of such decisions by those who know little or nothing about the measurement of writing competency. Armed with their knowledge and skills, faculty can contribute to the growing body of research in assessment. Finally, informed faculty can appropriately influence their own institutions to utilize direct measures of writing ability.

Works Cited

- Barber, Phyllis, A. Personal interview. 9 Apr. 1987.
- Barritt, Loren, Patricia L. Stock, and Francelia Clark. "Researching Practice: Evaluating Assessment Essays." College Composition and Communication 34 (1986): 315-327.
- Betz, Norman,. "Using Portfolios in Writing Assessment: Should It Replace or Supplement Exit Testing in Freshman Composition?" Conference on Composition and Communication Convention. Atlanta, Ga. 19 Mar. 1987.
- Betz, Renee T. "Difficulties in Developing Portfolio Assessment." Conference on Composition and Communication Convention. Atlanta, Ga. 19 Mar. 1987.
- Brown, Stuart C. "Scoring Standards for an Upper-Division Writing-Proficiency Test: A Description." Conference on Composition and Communication Convention. Atlanta, Ga. 21 Mar. 1987.
- Bruffee, Kenneth A. "Beginning a Testing Program: Making Lemonade." Writing Assessment: Issues and Strategies. Eds. Karen L. Greenberg, Harvey S. Wiener, and Richard A. Donovan. New York: Longman, 1986. 93-97.
- Cooper, Charles R. "Holistic Evaluation of Writing." Evaluating Writing: Describing, Measuring, Judging. Eds. Charles R. Cooper and Lee Odell. Urbana, Ill.: NCTE, 1977. 3-31
- Diederich, Paul B. Measuring Growth in English. Urbana, ILL.: NCTE, 1974.
- Fowles, Mary E. "The Development of Quality-Assurance Standards in Essay Testing." Conference on Composition and Communication Convention. Atlanta, Ga. 19 Mar. 1987.
- Gere, Anne Ruggles. "Written Composition: Toward a Theory of Evaluation." College English 42 (1980): 44-58.
- Glick, Gretchen, and Barbara Cole. "Matching Assessment Purposes and Instruments." Notes From the National Testing Network in Writing 6 (1986): 5-6.
- Greenberg, Karen L. "Defining, Teaching, and Testing Writing: The Teacher's Perspective." Conference on Composition and Communication Convention. Atlanta, Ga. 19 Mar. 1987.
- . "Competency Testing: What Role Should Teachers of Composition Play?" College Composition and Communication 33 (1982): 366-376.
- Lloyd-Jones, Richard. "Primary Trait Scoring." Evaluating Writing: Describing Measuring, Judging. Eds. Charles R. Cooper and Lee Odell. Urbana, Ill.: NCTE, 1977. 33-66

- Matalene, Carolyn B. "Objective Testing: Politics, Problems, Possibilities." College English 44 (1982): 368-381.
- Moran, Michael G., and Ronald F. Lunsford, eds. Research in Composition and Rhetoric. Westport, Conn.: Greenwood Press, 1984.
- Odell, Lee. "Defining and Assessing Competency in Writing." The Nature and Measurement of Competency in English. Ed. Charles R. Cooper. Urbana, Ill.: NCTE, 1981. 95-138.
- Odell, Lee, and Charles R. Cooper. "Procedure for Evaluating Writing: Assumptions and Needed Research." College English 42 (1980): 35-43.
- Raymond, James C. "What We Don't Know about the Evaluation of Writing." College Composition and Communication 33 (1982): 399-403.
- White, Edward M. "The Centrality of Measurement." Notes From the National Testing Network in Writing 7 (1986): 3.
- . "Holisticism." College Composition and Communication 35 (1984): 400-409.
- . Teaching and Assessing Writing. San Francisco: Jossey-Bass Publishers, 1985.